Chapter-5
5.0. Materials

5.1. Hardware Components
In present work all the calculations were carried out with high frequency computational analysis such as molecular modeling, energy minimizations, design and optimization of lead molecules, protein ligand interaction studies by molecular docking etc., a Hi-end server (Pentium IV 3.4 MHzs, AMD Athlon 64 bit, Dual processor with 1GB RAM) manufactured by HCL Corporation, Pondicherry, India was used.

5.2. Software Components
Most of the softwares used were either Windows or Linux platform based which were well accepted and referred in various publications at high rated research journals. Academic license was obtained for the commercial software used in the present study by requesting the concerned suppliers. The software used in the present study was briefly detailed below.

5.2.1. MODELLER
MODELLER is a computer program used in producing homology models of protein tertiary structures as well as quaternary structures which implements a technique inspired by nuclear magnetic resonance known as satisfaction of spatial restraints, by which a set of geometrical criteria are used to create a probability density function for the location of each atom in the protein. The method relies on an input sequence alignment between the target amino acid sequence to be modeled and a template protein whose structure has been solved. MODELLER was originally written and is currently maintained by Andrej Sali at the University of California, San Francisco. Although it is freely available for academic use, graphical user interfaces and commercial versions are distributed by Accelrys (http://www.salilab.org/modeller).

5.2.2. PyMOL
PyMOL is an open-source, user-sponsored, molecular visualization system, which is well suited to produce high quality 3D images of small molecules and biological macromolecules such as proteins. According to the author, almost a quarter of all published images of 3D protein structures in the scientific literature were made using PyMOL. (http://www.delanoscientific.com/)

5.2.3. Visual molecular dynamics (VMD)
VMD is a molecular modelling and visualization computer program which is
primarily developed as a tool for viewing and analyzing the results of molecular dynamics simulations, but it also includes tools for working with volumetric data, sequence data, and arbitrary graphics objects. Molecular scenes can be exported to external rendering tools such as POV-Ray, Renderman, Tachyon, VRML, and many others. (http://www.ks.uiuc.edu/Research/vmd/)

5.2.4. AutoDOCK-Tool
AutoDock is a suite of automated docking tools designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure. AutoDock actually consists of two main programs: AutoDock performs the docking of the ligand to a set of grids describing the target protein; AutoGrid pre-calculates these grids. In addition to using them for docking, the atomic affinity grids can be visualised. This can help, for example, to guide organic synthetic chemists design better binders. (http://autodock.scripps.edu/).

5.2.5. Clustal
Clustal is a widely used multiple sequence alignment computer program. The latest version is 1.83. There are two main variations:

- Clustal-W: Available on line at EMBL server
- Clustal-X: This version has a graphical user interface. It is available for Windows, Mac OS and Unix/Linux.

5.2.6. Hyperchem
HyperChem is the software for molecular modeling, energy minimization and simulation of lead and drug molecules. It calculates the QSAR properties of small existing molecules in database. It is windows based commercial software. http://www.hyper.com/.

5.2.7. Chem Office ultra 7.0
ChemOffice Ultra is the ultimate chemistry and biology suite designed for chemists, which is a suite of software consisting of ChemDraw, Chem3D and ChemFinder. It has encyclopedia of chemical structures, drugs and biological properties with over 10,000 monographs on single substances or groups of related compounds.

5.3. On line Tools
In addition to the above software, various on line computational tools used in the present study were as denoted below.
5.3.1. National Center for Biotechnology Information (NCBI)
The NCBI is part of the United States National Library of Medicine (NLM), a branch of the NIH, located in Bethesda, Maryland and was founded in 1988. The NCBI houses genome sequencing data in GenBank and an index of biomedical research articles in PubMed Central and PubMed, as well as other information relevant to biotechnology. All these databases are available online through the Entrez search engine. (www.ncbi.nlm.nih.gov).

5.3.2. Protein Data Bank (PDB)
The Protein Data Bank (PDB) is a repository for 3-D structural data of proteins and nucleic acids. These data, typically obtained by X-ray crystallography or NMR spectroscopy, are submitted by biologists and biochemists from around the world, are released into the public domain, and can be accessed for free. The mission of the PDB is to maintain a single Protein Data Bank Archive of macromolecular structural data that is freely and publicly available to the global community. (www.rcsb.org/pdb).

5.3.3. Pfam
Pfam is a collection of protein motifs and families maintained by the Bioinformatics group at the Sanger Centre. Pfam hidden Markov models (HMMs) and the Prosite generalized profiles were developed based on distinct theoretical backgrounds. (http://www.sanger.ac.uk/Software/Pfam/).

5.3.4. SWISS-Prot (Expasy)
This server is manually curated biological database of protein sequences created in 1986 by the Swiss Institute of Bioinformatics and the European Bioinformatics Institute. Swiss-Prot strives to provide reliable protein sequences associated with a high level of annotation (such as the description of the function of a protein, its domains structure, post-translational modifications, variants, etc.), a minimal level of redundancy and high level of integration with other databases. (http://expasy.org/sprot/). This program is available from European Bioinformatics Institute ftp server (http://www.ebi.ac.uk/Tools/clustalw/index.html)

5.3.5. Molinspiration Server
JME Molecular Editor is a Java applet which allows to draw / edit molecules and reactions (including generation of substructure queries) and to depict molecules directly within an HTML page. Editor can generate Daylight SMILES or MDL mol file of created structures. Due to many requests, the applet (in form of a jar file)
has been released to the public and become a standard for molecular structure input on the web with more than 6500 installations worldwide. As recognition of this generous gesture, Molinspiration provides this space for the JME Home. Molinspiration can also offer help with installation and deployment of the JME. (http://www.molinspiration.com/docu/webme/)

5.3.6. Jmol First Glance
First Glance in Jmol is the easiest way to look at the 3D structures of proteins, DNA, RNA, and their complexes. It is in use by Nature Structural and Molecular Biology and the ConSurf Server (which automatically colors amino acids by evolutionary conservation. (http://molvis.sdsc.edu/fgij/index.htm)

5.3.7. PDBSUM
The PDBsum is a pictorial database that provides an at-a-glance overview of the contents of each 3D structure deposited in the Protein Data Bank (PDB). This server provides cleft and groves on the surface of protein molecules deposited in protein data bank. Pdbsum can also give ligand binding site with Ligplot graphs in two dimensional appearances (Laskowski et al., 2005). (http://www.ebi.ac.uk/pdbsum/)

5.3.8. PRODRG
The PRODRG will convert coordinates of small molecules in PDB format to various formats of GROMACS, GROMOS, WHAT IF, CNS, AUTODOCK 2.4 and AUTODOCK 3.0 etc. The output files can be used for the further analysis in various softwares. In addition coordinates for hydrogen atoms are generated. (www.davapc1.bioch.dungee.ac.uk.)

5.3.9. WHAT IF
What if (Vriend, 1990) is a web interface server for structural validation, build, repair, residue, protein and docking analyser. It accepts Protein Data Bank (PDB) formatted files or PDB accession numbers as input and calculates, identifies, 2D and 3D graphs, reports and/ or evaluate it. What if web server is freely accessible at http://swift.cmbi.ru.nl/servers/html/index.html

5.3.10. ProSA-web
ProSA program (Wiederstein et al., 2007) which exploits the advantages of interactive web-based applications for the display of scores and energy plots that highlight potential problems spotted in protein structures. In particular, the quality scores of a protein are displayed in the context of all known protein structures and
problematic parts of a structure are shown and highlighted in a 3D molecule viewer. The service specifically addresses the needs encountered in the validation of protein structures obtained from X-ray analysis, NMR spectroscopy and theoretical calculations. (http://prosa.services.came.sbg.ac.at)

5.4. Sequence analysis

The most important point in any homology modeling study, besides the choice of the reference, is the alignment of the sequences. The greatest attention was thus paid to the careful construction of a robust alignment. Various types sequence analysis was carried out through retrieving the sequences either from NCBI or SWISS-Prot databases. Sequence homology search was conducted through the blast-P program available at NCBI. Homology modeling of target sequence needs a template crystal structure coordinates which were obtained by performing blast-P at NCBI with selection of database as PDB at (www.ncbi.nlm.nih.gov/blast.html.) (Altschul et. al., 1990). The coordinates of selected crystal structures of sequence similar structures of target protein were obtained from PDB and used for prediction of 3D structure of target protein using MODELLER 9v6. In order to identify conserved and variable regions of the sequences and in determining the most robust gap arrangement, multiple sequence alignment of all homologous proteins of the target sequence Clustal-W (Chenna et. al., 2003) with appropriate parameters were used as per the specified instructions. The Clustal-W alignment file of the selected sequences was used for the basic parameter for further creating the phylogenetic tree with target, as query sequence.

5.5. Phylogenetic analysis

Sequences were aligned using Clustal-X (Thompson et al., 1997), Clustal performs a global-multiple sequence alignment by the progressive method. The steps include:

a) Perform pair-wise alignment of all the sequences by dynamic programming
b) Use the alignment scores to produce a phylogenetic tree by neighbor-joining
c) Align the multiple sequences sequentially, guided by the phylogenetic tree

Thus, the most closely related sequences are aligned first, and then additional sequences and groups of sequences are added, guided by the initial alignments to produce a multiple sequence alignment showing in each column the sequence variations among the sequences. Sequence contributions to the multiple
sequence alignment are weighted according to their relationships on the predicted evolutionary tree. Weights are based on the distance of each sequence from the root. The alignment scores between two positions of the multiple sequence alignment are then calculated using the resulting weights as multiplication factors. As more sequences are added to the profile, gaps accumulate and influence the alignment of further sequences. Clustal calculates gaps in a novel way designed to place them between conserved domains. Gaps found in the initial alignments remain fixed. New gaps are then introduced into the multiple alignment when more sequences are added, but gaps can never be deleted, only added. Clustal also implements methods, which try to compensate for the scoring matrix (e.g., PAM), expected number of gaps, and differences in sequence length.

5.6. Homology modeling

The 3-D homology models of given target protein sequence was predicted using crystal structural coordinates of templates on the basis of sequence alignment. All steps of homology modeling and refinement were carried out through MODELLER 9v6 using base line commands specified by software supplier (Sali and Blundell, 1993). The method described below is used in the present study to predict the 3D models of Alr.Ddl and DltA enzymes.

5.6.1. Preparation of input files for MODELLER

There are three kinds of input files are required to perform homology modeling using MODELLER. They are PDB atom files with coordinates for the templates, the alignment file with alignment of the template structures with the target sequence, and finally .PY file, a MODELLER command file that instructs MODELLER what to do.

5.6.1.1. Atom file

Each atom file is named as code.pdb where code is a short protein code, preferably the PDB code. The atom file contains the only protein co-ordinates without hetero atoms while modeling target protein.

5.6.1.2. Alignment file

One of the formats for the alignment file (Figure: 5.1) is related to the PIR data base format which is the preferred format for homology modeling by MODELLER.

5.6.1.3. Script file

The .PY file contains commands for MODELLER. A sample steering file is to produce one model of sequence. A number of intermediary files were created as
the program proceeds. After 10 minutes, the final protein model is written to file protein.B999901. A log file was also created with information about the run.

Fig 5.1. Alignment file with query sequence and template

5.6.2. Flowchart of Homology Modeling by Modeller

This section describes a flow chart of homology modeling by MODELLER, as implemented in the ‘model’ .PY script which also be used for variety of modeling tasks not only for comparative modeling. Input: script file (steering file; alignment file, PDB file(s) for template(s).

Output:
- .log long file
- .ini initial conformation for optimization
- .rsr restraints file
- .sch VTEM schedule file
- .B999???? PDB atom file(s) for the model(s) of the target sequence
- .V9999???? Violation profiles for the model(s)

The main MODELLER routines used in each step are given in parentheses.

1. Read and check the alignment between the target sequence and the template structures (READ_ALIGNMENT and CHECK_ALIGNMENT).
2. Calculate restraints on the target from its alignment with the templates:
   a. Generate molecular topology for the target sequence (GENERATE_TOPOLOGY). Disulfides in the target are assigned here from equivalent disulfides in the templates (PATCH_DISULFIDES). Any user defined patches are also done here (as defined in Top routine 'special patches').
b. Calculate coordinates for atom that have equivalent atoms in the templates as an average over all templates (TRANSFER_XYZ) (alternatively, read the initial coordinates from a file).

c. Build the remaining unknown coordinates using internal coordinates from the charm topology library (BUILD_MODEL).

d. Write the initial model to a file with extension .ini (WRITE_MODEL).

e. Generate stereochemical, homology derived, and special restraints (MAKE_RESTRAINTS) (alternatively, skip this and assume the restraints file already exists).

f. Write all restraints to file with extension .rst (WRITE_RESTRAINTS).

3. Calculate model(s) that satisfy the restraints as well as possible, for each model:

a. Generate the optimization schedule for the variable target function method (VTFM)

b. Read the initial model (usually from the .ini file from 2.d) (READ_MODEL)

c. Randomize the initial structure by adding a random number between DEVIATION angstroms to all atomic positions (RANDOMIZE_XYZ).
d. optimize the model

Partially optimize the model by VTFM; Repeat the following steps as many
time specified by optimization schedule:
Read all restraints by rd_restraints (READ_RESTRANTS)
Select only the restraints the operate on the atoms that are close enough
in sequence, as specified by the current step of VTFM
(PICK_RESTRANTS).
Optimize the model by conjugate gradients, using only currently selected
restraints (OPTIMIZE).
Refine the model by simulated annealing with molecular dynamics, if so selected:
  ➢ Do a short conjugate gradient optimization (OPTIMIZE).
  ➢ Increase temperature in several steps and do molecular dynamics
    optimization at each temperature (OPTIMIZE).
  ➢ Decrease temperature in several steps and do molecular
    dynamics optimization at each temperature (OPTIMIZE)
  ➢ Do a short conjugate gradient optimization (OPTIMIZE).
e. Calculate the remaining restraints violations and write them out (ENERGY).
f. Write out the final model to a file with extension .B99999???? Where????
Indicates the model number (WRITE_MODELS). Also write out the violation
profile. Also write superposed templates and model, if so selected by
FINAL_MALIGN3D = 1.

5.7. Evaluation of the built 3D protein model

A protein 3D model derived from homology modeling technique may have some
sources of errors. It is important, therefore, to have an assessment of structure’s
quality and to be able to identify regions that may need modifications especially at
protein folding and turns. The aim of model evaluation is to determine whether
the built model is acceptable and suitable to use for molecular analysis such as
docking and dynamics. The accuracy of the comparative built structures were
tested using the ENERGY command of the MODELLER program (Sali and
Blundell, 1993) and tools like PROCHECK, (Laskowski et al., 1993) and WHAT IF
(Vriend, 1990) In addition, the variability of the homology model has been
compared by superposition of Ca traces and backbone atoms model and crystal
structures, from which the RMSD value for positional differences between
equivalent atoms calculated with SPDV (Guex, 1999). which clearly judges the
accuracy of model.

5.8. PROCHECK
The PROCHECK suite of programs provides a detailed check on the stereochemistry of a protein structure. The stereo chemical parameter checks implemented in PROCHECK are derived from high-resolution protein structures, against which the structure is compared on a residue-by-residue basis. The criteria are Ramachandran plot, peptide bond planarity, C-alpha tetrahedral distortion, non bonded interactions, hydrogen bond energies, and closeness of side chain dihedral angles to ideal values.

5.9. Prediction of secondary structure of protein
The prediction of protein secondary structure is a major parrot of the general protein folding problem and the method of obtaining some structural information for any sequence. Secondary structure predication is important in establishing alignments during homology modeling. Secondary structure analysis is carried out through the ProFunc and PDBSUM server (Laskowski et. al., 2005), which gives the clear data of protein, alpha helices, sheets, turns, beta hairpins, beta bulges, gamma turns etc.,

5.10. Establishment of motif or domains structure of protein model
The motifs or domains of given 3-D model was obtained through online tool at MotifScan, Pfam server, which provides the different motifs and conserved regions that present in given protein structure.

5.11. Determination of active site residues
Determination of active site amino acid residues of given protein was performed with the help of literature survey from wet lab results. Based on high identity with active site residues from the crystal structure, the residues of active site of target protein have been established perfectly. The crystal structures are submitted to PDBSUM server which provides the catalytic sites and from that one can determine the conserved and catalytic residues in the active site of the built protein model by sequence alignment. In our study we have identified active site from the literature and those residues were investigated in homology model with visualization tool like Pymol.

5.12. Performance of salt bridge analysis of protein models
Salt bridge analysis of modeled protein was carried out through the online server of Jmol First glance (http://firstglance.jmol.org/) from which the internal ionic
interaction of the protein can be best established.

5.13. Molecular dynamics setup

Molecular dynamics (MD) is a computational method that calculates the time dependent behavior of a molecular system. MD simulations provide detailed information on the fluctuations and conformational changes of proteins and nucleic acids, and they are now routinely used to investigate the structure, dynamics and thermodynamics of biological molecules and their complexes. The basic idea of molecular dynamics (MD) is to study atomic fluctuations in solvated system. The principle behind this is application of classical Newton’s equation. All simulations reported in this thesis were performed with GROMACS molecular simulation package (Berendsen et al., 1995). Before MD, the ionization states of amino acids were set to mimic a neutral PH environment i.e. all Lys and Arg carried net positive charge, and all glutamic acid and Aspartic acid carried a net negative charges. The histidine residues were in doubly protonated condition. A random generation of 100 models from the starting structure was calculated and subsequently the best model with the low RMS value of superposition using Swiss-pdb viewer (Guex et al., 1999) was subjected for further analysis. The best modeled protein were solvated with water molecules in a truncated octahedron box. The size of the box was set to 0.9 nm distance from the surface of the protein. The Single Point Charge (SPC) water model (Berendsen et al., 1987) and ions (Na+ and Cl −) was built. The box model, first with explicit water and then with ions was added to protein containing truncated octahedron box, this was submitted to 400 steps of energy minimization using the steepest descent algorithm till an energy gradient was reached and it was found to be the most appropriate energy gradient to relax the models and afford well Ramachandran plots. In order to constrain all bond length in protein, the LINCS (Hess et al., 1997) algorithm was used. For water molecule bond length constrain, the SETTLE algorithm was implemented (Miyamoto and Kollman, 1992). The electrostatic and Van Der Waal forces are implemented using particle mesh Ewald potential method (Essmann et al., 1995) and Lennard-Jones potential method respectively. All full MD simulations were performed at 500ps with no restriction using two fs of integration time, constant temperature and pressure. The temperatures of the proteins and solvent molecules were each coupled separately, using (V-resacle) Berendsen thermostat algorithm (Berendsen et al., 1984). The pressure was
coupled using (Parrinello-Rahman) Berendsen algorithm at 1 bar with coupling constant $T_p = 1$ ps. Co-ordinates and energy terms (total, kinetic and potential for the whole system and electrostatic, distance dependent, distance-independent reaction force field) were saved for each ps. With the aim of evaluating the system stabilization throughout the molecular dynamics time, the total, kinetic and potential energy was plotted versus time. The stabilization was assessed by graphics visualization using VMD (Humphrey et al., 1996) and Xmgrace.

5.13.1. Atomic force Fields parameterization

The empirically derived potential function that describes interactions present between the atoms, in a molecule or separate molecules, is usually called as force field. The force field applied for calculations of potential interaction during MD simulations is GROMOS 96 (van Gunsteren et al., 1996). In GROMOS 96 force field, the non-bonded interactions are a sum of electrostatic and van der Waal contributions.

Non-bonded interaction empirical calculation is defined as follows.

$$ U_{nb} = \sum_{\text{atom pairs}} \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{ij}} + \sum_{\text{atom pairs}} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) $$

Where $q_i$ is the partial charges present on atoms $i$ and $j$, $r_{ij}$ is the distance between the two atoms to be under consideration, $\varepsilon_0$ denotes electric permittivity of simulated system, and $A$ and $B$ are the Lennard-Jones parameters that purely dependent on chemical nature of interacting atoms.

Bonded interaction

$$ U_{bn} = \sum_{\text{bonds}} k_b (r_{ij} - r_0)^2 + \sum_{\text{angles}} k_\theta (\theta_{ijk} - \theta_0)^2 + \sum_{\text{tumors}} k_\phi [1 + \cos(n\phi_{ijkl} - \delta)] $$

$k_b$, $k_\theta$ and $k_\phi$ are force constants respectively, $r_0$ is the equilibrium bond length and $\theta_0$ the equilibrium angle. $\theta$ is the angle between atoms $i$, $j$ and $k$. $n$ is the number of minima per full turn of the torsional angle $\phi$, and $\delta$ is the location of the first barrier.

5.13.2. Atomic flexibility analysis

The inherent flexibility of amino acids during MD simulations is measured using Root Mean Square Fluctuation (RMSF) term. The calculation of RMSF at time $t$ of atoms in a molecule with respect to static structure is defined as
RMSF(t) = \left[ \frac{1}{N} \sum_{i=1}^{N} [r_i(t) - \langle r_i \rangle]^2 \right]^{\frac{1}{2}}

Where \( r_i(t) \) is the position of atom type \( i \) at time \( t \) and \( N \) is the number of atoms. With the use of above equations which is readily available in GROMACS program, the significant movement of ET residues was calculated over stipulated MD simulations period. \( \langle r_i \rangle \) is the average position of atom \( i \) in MD simulations.

5.13.3. Molecular dynamic simulation

After successful completion of molecular dynamic simulation set up and force field parameterization, MD simulations are to be implemented. In present study, Leap-frog version of Verlet algorithm was exploited for generating time averaged structural conformations with respect to forces that act on individual atoms (van Gunsteren and Berendsen, 1988). Leap-frog version of Verlet algorithm uses Newton's second law of motion.

\[ F_i = \frac{\partial U(r_1, \ldots, r_N)}{\partial r_i} \]

Newton's law of motion is thus used in MD simulations to calculate the forces up on successive configurations of the system. The velocities, constant temperatures and pressure were constantly maintained. The MD (Molecular Dynamic) simulation algorithm used in present work is implemented in GROMACS package. The system temperature and pressure were kept constant throughout the MD simulation period.

5.13.4. Analysis of Molecular dynamic results

MD simulations produce bunch of structural conformations at different time scale. Therefore, well planned analysis of bunch of structural conformations can provide vital clues of molecular function exactly. From all simulations generated from starting experimental model the Root Mean Square Fluctuation (RMSF), Root mean square deviation (RMSD), potential, kinetic and total energies were analyzed. The stabilities of intramolecular hydrophobic interactions were evaluated in terms of Lennard-Jones potential. Lennard-Jones potential is a good approximation of Van Der Waal (VDW) stabilization energies.

5.14. Design and selection of Ligand molecules

103
In this scenario of designing ligands, all the modifications were done by taking into consideration of a database of substituents and spacers (linkers) obtained by substructure analysis of a collection of current drugs, development of drugs and other molecules with biological activity containing about 17000 entries developed in Molinspiration server (http://www.molinspiration.com). Nearly 100 lead molecules were designed based on these considerations, by using HyperChem 7.5% Professional, which is a molecular modeling and simulation software that lets to perform complex chemical calculations. For designing each of the ligand a number of rotatable bonds were assigned manually, letting the side chains rotate and the keeping the backbone rigid. It was taken into consideration during the design of ligands, that the number of bond rotations was kept limited and to one side of the molecule. In addition to assigning rotations the unpolar hydrogens were removed and the partial charges from these were added to the carbon that held the hydrogen. All these preparations were done to each ligand using HyperChem 7.5. These 100 lead molecules were then subjected to Chem Office Ultra 7.0 (Muegge, 2003). to find their log P values and these molecules were given chemical formulae names by using the same software. All these molecules were then analyzed for their ability to follow Lipinski's Rule-of-five (Lipinski, 2001) by subjecting them to Molinspiration server (http://www.molinspiration.com). The rule describes molecular properties important for drug's pharmacokinetics in the human body, including their absorption, distribution, metabolism and excretion (ADME).

The Lipinski's Rule of five states that in general orally active drug should has

- Not more than 5 hydrogen bond donors (OH and NH groups)
- Not more than 10 hydrogen bond acceptors (notably N and O)
- A molecular weight under 500g/mol
- A partition coefficient log P less than 5.
- Topological Surface area (TPSA)

Among all the designed leads the molecules of high ranking which follow Lipinski's rule were selected and further analyzed for binding with the protein model using docking tools.

5.14.1. Preparation of files for AUTODOCK

The advanced molecular docking program AutoDock 4.0 which uses a powerful Lamarckian genetic algorithm (LGA) method for conformational search and
docking, was applied for the automated molecular docking simulations. Briefly, the LGA described the relationship between the antagonists and receptors by the translation, orientation, and conformation of the antagonists. These so-called 'state variables' were the ligands' genotype, and the intramolecular energies were the antagonists' phenotype. The environmental adaptation of the phenotype was reverse transcribed into its genotype and became heritable traits. Each docking cycle or generation, consisted of regimen of fitness evaluation, crossover, mutation, and selection. A Solis and Wets local search (Ruth Huey et al, 2008) was carried out to the energy minimization on a user-specified proportion of the population. The docked structures of the ligands were generated after a reasonable number of evaluations. The whole docking scheme could be stated as follows.

First, the receptor molecules were checked for polar hydrogen and assigned for partial atomic charges, the PDBQT file was created, and the atomic salvation parameters were also assigned for the macromolecules. Meanwhile, all of the torsion angles of the antagonists that would be explored during molecular docking stage were defined. Therefore, it allowed the conformation search for ligands during molecular docking process.

Second, the 3D grid was created by Auto Grid algorithm (Ruth Huey et al, 2008) to evaluate the binding energies between the antagonists and receptors. In this stage Air, Ddi and DltA substrates, inhibitors and design lead molecules receptor was embedded in the 3D grid and probe atom was placed at each grid point. The affinity and electrostatic potential grid were calculated for varies type of atoms in the ligands. The energetic configuration of a particular ligand was found by trilinear interpolation of affinity values and electrostatic interaction of the eight grid points around each atom of the ligand. Third, a series of the docking parameters were set on. The atom types, generations and run numbers for LGA algorithm were properly assigned according to the requirement of the Amber force field. The number of generations, energy evolutions, and docking runs were set to 25,00,000, and 27,000, respectively. The kind of atomic charges were assigned as computes Gasteiger charges for air, ddi and DltA receptor and Gasteiger-Marsili (Gasteiger et al., 2003) for the ligands.
5.14.2. Protein – Lead molecules binding studies using AutoDock Tool
Preparation of the target protein with the AutoDockTools software involved the addition of polar hydrogen atoms to the macromolecule, a necessary step for the correct calculation of partial atomic charges. Gasteiger charges are calculated for each atom of the macromolecule in AutoDock 4.0 instead of Kollman charges, which were used in the previous versions of this program. AutoDock4.0 distinguishes atoms that can accept hydrogen bonds (NA, OA, and SA) and those unable to do so (N and S). It also distinguishes hydrogen atoms that can participate in a hydrogen bond (HD) and non-hydrogen-bonding hydrogen atoms (H). Additionally, an electrostatic map and a desolvation map were calculated. Ligand pdbq file was obtained from PRODRG2 Server (Schuettelkopf and Van Aalten, 2004). All the atom types were checked in the ligand and modified when needed. Ligands were chosen charged and hydrogens were added in order to fill all empty valences, and the Gasteiger charges were calculated for the atoms, and were then saved. In order to run AutoDock, grid maps have to be calculated. This was done by using the module AutoGrid, for each of the ligand and with the same parameters: number of grid points in X, Y and Z were taken as 60X60X60 (this covers the active site extensively and let the ligand move without any constraints regarding the box size), spacing between grid points, 0.375 Å and a common grid centre. The grid centre was chosen slightly of the centre axis of the active site in order to avoid any symmetry problems that might arise. Docking Protocol. AutoDock 4.0 was used for the docking simulation. We employed the Lamarckian genetic algorithm (LGA) for ligand conformational searching. LGA is a hybrid of a genetic algorithm and a local search algorithm. This algorithm first builds a population of individuals (genes), each gene being a different random conformation of the docked compound. The population undergoes simulated evolutionary development with processes of phenotypic mapping, fitness evaluation, natural selection, crossover, and elitist selection occurring in each generation. The local search algorithm then performs energy minimizations on a user-specified proportion of the population of individuals. Individuals with the lowest resulting energies are then transferred to the next generation, and the process is repeated. We set important docking parameters for the LGA as follows: population size of 150 individuals, maximum of 2.5 million energy evaluations, maximum of 27 000 generations, one top individual to survive to the next...
CHAPTER 5 MATERIALS & METHODS

generation automatically, mutation rate of 0.02, crossover rate of 0.8, 100 docking runs, and random initial positions and conformations. The probability of performing a local search on an individual in the population was set to 0.06, and the maximum number of iterations per local search was set to 150. These standardized docking parameters create a file for each ligand and hence AutoDock program for each ligand was run. Each docking job produced 100 docked conformations, which were clustered by the root-mean-square deviation (rmsd), which measured the difference between the coordinates of their constituent atoms, so that conformations in each cluster differed by less than 2.0Å.

<table>
<thead>
<tr>
<th>Clus</th>
<th>Lowest</th>
<th>Run</th>
<th>Mean</th>
<th>Num</th>
<th>Histogram</th>
</tr>
</thead>
<tbody>
<tr>
<td>-ter</td>
<td>Docked</td>
<td>Docked</td>
<td>Energy</td>
<td>Energy</td>
<td>Clus</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-10.59</td>
<td>44</td>
<td>-10.59</td>
<td>1</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>-8.94</td>
<td>66</td>
<td>-8.94</td>
<td>1</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>-8.63</td>
<td>45</td>
<td>-8.42</td>
<td>2</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>-8.42</td>
<td>19</td>
<td>-7.96</td>
<td>5</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>-8.39</td>
<td>33</td>
<td>-7.83</td>
<td>3</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>-8.24</td>
<td>6</td>
<td>-7.55</td>
<td>18</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>-7.85</td>
<td>48</td>
<td>-7.57</td>
<td>7</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>-7.80</td>
<td>7</td>
<td>-7.79</td>
<td>2</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>-7.67</td>
<td>32</td>
<td>-7.44</td>
<td>3</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>-7.66</td>
<td>4</td>
<td>-7.50</td>
<td>2</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>-7.59</td>
<td>9</td>
<td>-7.59</td>
<td>1</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>-7.57</td>
<td>39</td>
<td>-7.57</td>
<td>1</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>-7.33</td>
<td>8</td>
<td>-7.08</td>
<td>2</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>-7.33</td>
<td>34</td>
<td>-7.33</td>
<td>1</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>-6.61</td>
<td>13</td>
<td>-6.61</td>
<td>1</td>
<td>#</td>
</tr>
</tbody>
</table>

Basing on the dig file the best docking interactions of the ligand were observed by using PMV viewer 1.5.4 (http://autodok.scripps.edu/). Analysis of docking interaction gives better picture of the amino acids involved in the binding of ligand molecules. Based on this we can identify the amino acids involved in the active site.